

Space reconstruction by primary visual cortex activity: a parallel, non-computational mechanism of object representation

Moshe Gur

Department of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa, Israel

The current view posits that objects, despite changes in appearance, are uniquely encoded by ‘expert’ cells. This view is untenable. First, even if cell ensemble responses are invariant and unique, we are consciously aware of all of the objects’ details. Second, in addition to detail preservation, data show that the current hypothesis fails to account for uniqueness and invariance. I present an alternative view whereby objects’ representation and recognition are based on parallel representation of space by primary visual cortex (V1) responses. Information necessary for invariance and other attributes is handled in series by other cortical areas through integration, interpolation, and hierarchical convergence. The parallel and serial mechanisms combine to enable our flexible space perception. Only in this alternative view is conscious perception consistent with the underlying architecture.

‘Where there is a question of “mind” the nervous system does not integrate itself by centralization upon one pontifical cell. Rather it elaborates a million-fold democracy whose each unit is a cell’ [1].

It is time for a different view on the neural basis of conscious object perception

How the retinal image is transformed into our object-based 3D perception has been the focus of much research in the past five decades. While most investigators have shied away from dealing with the question of how objective, physical brain activity generates private subjective percepts, trying to understand what brain activity is likely to generate conscious visual perception has attracted much attention recently [2–4]. Let me stress that here I do not deal with the question of how neural activity generates conscious perception; rather, I discuss the neuronal substrate of perception: what neural activity results in perception? This question is relevant to all models wishing to find the relationship, or

correlation, between activity in various visual areas and conscious perception. Models that aim to describe a path from visual input to action that bypasses perception are not considered.

Practically all models trying to explain space perception are influenced by the notion that the ‘brain is a remarkable computer’ [5]. Thus, all suggested brain mechanisms are such that can be, in principle, instantiated by computer programs. This approach is powerful in allowing rigorous simulation and testing of various models but, as I show below, it restricts our thinking to ‘computer-friendly’ theories that ignore *prima facie* perceptual evidence. Here I offer a different, non-computational view that is driven by well-accepted anatomical, physiological, and perceptual data (Box 1).

Current view of object representation and recognition

The dominant view is that images are analyzed into edges and line segments by V1 feature-selective cells and that, after several steps of hierarchical convergence and integration, small ensembles of expert cells, by their collective responses, represent objects uniquely and invariantly.

Glossary

- 2-Deoxyglucose (2DG) imaging:** uses the radioactive tracer ^{14}C -2-deoxy-D-glucose to image active (glucose-consuming) brain areas.
- Fovea:** a small retinal area of ~ 1.5 mm diameter with a high density of photoreceptors. Responsible for high-acuity vision.
- Hypercolumn:** a module in V1 comprising orientation and eye dominance columns. All cells within a hypercolumn receive input from the same VF area.
- Lateral geniculate nucleus (LGN):** a subcortical relay nucleus receiving its sensory input from the retina. Its main output is to V1.
- Receptive field (visual) (RF):** an area in the outside world or in the retina stimulation of which leads to responses of a particular visual neuron.
- Retinotopy:** a projection of visual input from the retina to a brain area that preserves retinal topography.
- Simple cell (in V1):** has an orientation-specific RF with nonoverlapping regions that are excited by either light increments or light decrements.
- Sparse population coding:** coding an item such as a face by a relatively small number of nerve cells.
- Visual dorsal system:** originates in V1 and continues along the cortical dorsal surface into the parietal cortex. Believed to be involved in non-conscious analysis of spatial locations, shapes, and orientations of objects, leading to reaching and grasping movements.
- Visual ventral system:** originates in V1 and continues along the ventral cortical surface into the temporal cortex. Assumed to lead to conscious perception and recognition of visual shapes and objects.

Corresponding author: Gur, M. (mogi@bm.technion.ac.il).

Keywords: vision; object representation; recognition; conscious perception; parallel processing.

0166-2236/

© 2015 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tins.2015.02.005>

Box 1. On the uniqueness of space perception

All sensory attributes of the world except space are not perceived as such but are transformed by the brain into new, internal-only constructs. For example, wavelengths are transformed into colors and air movements into sounds. When seeing colors, hearing sounds, or feeling pain we have no direct information on the nature of the energies in the outside world that are the source of these internal percepts. So when a tree falls in the forest and there is no one around, there is no sound, just air movement. This internal-only representation of sensory data is often used to show that the brain is not a passive, computer-like data processor but rather an organ creating a new, not externally observed reality. For obvious reasons this subjective reality has created difficulties in accepting as valid data first-person reports on sensory perception. If perception of blue exists only inside one's head how can personal reports serve as reliable, objective scientific data? Generations of psychophysicists have found clever ways around this problem, so today such reports are important components of any scientific investigation of sensory perception. Interestingly, the unique attributes of space perception that make it much less subjective than other senses have not been appreciated. Space, unlike other sensory information, is not perceived as a unique 'something else' structure but rather as a one-to-one correspondence between the internal percept and the layout of actual space elements. At the center of the visual field, we perceive not only what is out there, but with an amazing exactitude of details and topography. Thus, our space perception can be objectively verified in at least two important ways. One is that, for a given spatial pattern, say a collection of dots of various shapes and intensities, all observers' reports will be very similar; all will state that the rightmost dot is more circular and less bright than the one below and so on. This high degree of exactitude shared by all observers lends considerable credibility to subjective reports. The second verification of subjective reports is the ability to compare such reports with objective measurements by physical devices such as a photometer. Thus, all basic spatial characteristics reported by an observer (e.g., location, size, intensity) are completely verifiable by objective measurements.

We can thus conclude that, since perception of the organization and structure of basic spatial elements closely corresponds to physical reality, evidence from perception is essential to our understanding of neural mechanisms of space perception.

Hubel and Wiesel's evidence [6,7] of orientation-selective cells in the cat and monkey V1 established the basic tenets for all succeeding models of object representation and recognition. They suggested that feature selectivity in V1 cells is generated by hierarchical convergence of cells with concentric receptive fields (RFs) (see [Glossary](#)) in the subcortical lateral geniculate nucleus (LGN) and predicted that further convergence in areas downstream from V1 will enable encoding of increasingly complex features and simultaneously allow a considerable degree of invariance.

The next two areas downstream from the monkey V1, V2 and V4, show only a modest increase in feature selectivity [8–12] and it is the temporal cortex where cells that clearly respond to elaborate integrative features are found. Numerous studies showed that cells in the monkey inferotemporal (IT) cortex are selective for various complex shapes, including faces [13–21]. The hierarchical transformations leading to face-selective cells are paralleled by an increase in spatial integration from cells integrating over a few minutes of arc in V1 to cells at the pinnacle of the hierarchy responding within a large portion of the visual field (VF). Research in homologous areas in the human visual cortex is consistent with single-cell data from the monkey IT cortex [22–25]. Experimentalists invariably identify the temporal cortex as the

site of object representation and recognition [15–17,20–25]. Presumably, responses of orientation-selective V1 cells give rise to perception of short line segments [26] while perception of more complex objects such as eyes or faces is predicated on the responses of cells in areas that are higher on the hierarchical ladder such as V4 or the IT cortex [27]. It is thought that the collective properties of expert cells enable individual objects to be recognized despite changes in global parameters such as size or viewpoint [15].

Single-cell recordings also showed that receptive field (RF) properties can be modulated by top-down mechanisms. Context, experience, and, most notably, attention affect the responses of cells in V1 [28], V2 and V4 [29,30], and the IT cortex [31,32].

The physiological findings from all levels of the visual cortex that were taken as confirmation of object representation by sparse population coding ([Box 2](#)) have influenced practically all computational models of object representation and recognition. Physiology-based models, either those stressing feedforward mechanisms [33–35] or those adding top-down or lateral interactions related to attention, detail scrutiny, or awareness [36–38], have at their core a process of hierarchical convergence leading to feature elaboration. Other models that are only loosely patterned after visual cortex physiology, while emphasizing various aspects of network interactions, are also assuming that objects are ultimately represented by cell assemblies generated by hierarchical convergence [39–41] ([Box 2](#)).

Fundamental to the current view of object representation are two main processes: encoding and hierarchical convergence. It assumes that once parallel retinal space information reaches feature-selective V1 cells, it is transformed into a code carried by cells that, by hierarchical convergence, respond to increasingly elaborate features. Suggested codes may be relatively simple, whereby single cells' responses encode spatial features such as lines or faces, or more complex ones where dynamic ensembles from various levels of the visual hierarchy combine to encode the required spatial information. How such a code is decoded into our parallel, detailed space perception is usually not dealt with.

Problems with the current view

Most physiological research regarding object representation has concentrated on the properties of face-selective cells found in the IT cortex. Here I thus use findings and ideas related to 'face cells' as examples of the more general case of object representation and recognition.

The idea that an ensemble of a few hundred expert cells can respond uniquely to thousands of faces yet be invariant to many possible changes in their global appearance (e.g., size, contrast, position, tilt, viewpoint, noise, shadows; see [Figure 1](#)) is generally accepted, although it is hard to think of any realistic implementation. Indeed, no system, physiological or physiologically inspired, was ever tested with stimuli that come close to our perceptual ability and the available physiological evidence is mostly anecdotal and does not support true invariance (see below). Moreover, even if we accept that expert cell ensemble responses are able to uniquely

Download English Version:

<https://daneshyari.com/en/article/4354159>

Download Persian Version:

<https://daneshyari.com/article/4354159>

[Daneshyari.com](https://daneshyari.com)